Gender equality and sex differences in personality: evidence from a large, multi-national sample

Preprint · July 2018

DOI: 10.17605/OSF.IO/6NUHX

1 author:

Tim Kaiser
University of Salzburg

Gender equality and sex differences in personality: evidence from a large, multi-national sample

Corresponding author:

Tim Kaiser

University of Salzburg, Department of Psychology

Hellbrunnerstrasse 34

Salzburg, Austria

Tim.Kaiser@sbg.ac.at

Abstract

Aim: findings on sex differences in personality are robust and stable across countries. Several studies found that these differences increase with greater societal gender equality. However, these studies have shortcomings as they either (a) did not use valid indicators of gender equality and development, (b) only studied broad domains of personality, (c) did not address issues of measurement invariance. The aim of this study is to replicate previous findings on the correlation between gender equality and sex differences in a methodologically robust way.

Method: a large, multinational (N = 926,383) dataset was used to examine sex differences in Big Five facet scores for 70 countries. Difference scores were aggregated to a multivariate effect size (Mahalanobis' *D*).

Results: effect sizes were large (average *D* = 1.84), but varied across countries. Countries' difference scores were related to an index of gender equality, revealing a positive weighted correlation of r = .394, p = .001.

Conclusion: using multivariate effect sizes derived from latent scores with invariance constraints, the study of sex differences in personality becomes more robust und replicable. Sex differences in personality should not be interpreted as results of unequal treatment, but as indicator of successful gender equality policies.

Keywords: personality, gender equality, sex differences, measurement

Introduction

Since the beginning of psychometric personality assessment, sex differences in personality inventories have been reported. Meta-analytic findings on these differences on normative data sets and the empirical literature reported by Feingold in 1994 (Feingold, 1994) suggested that these differences are robust and constant across age, year of data collection, education and nations. A large multinational later challenged these findings (P. Costa Jr., Terracciano, & McCrae, 2001). While essentially replicating the magnitude and robustness of sex differences in personality, the authors found larger sex differences in western cultures with higher gender equality. The authors concluded that these findings contradict sociological, gender role-based explanations for sex differences. Other predictors of larger sex differences in personality were proposed by Schmitt, Realo, Vorace and Allik (2009). In a sample of subjects from 55 nations, the authors reported larger sex differences in more developed countries. It was hypothesized that these findings can be explained by the fact that there is a general biological trend to stronger sexual dimorphism in environments that are rich in resources. As time progressed, the samples of studies adding to these finding became larger. Lippa (2010) reported a negative correlation between UN gender development indices and sex differences in the personality trait “Agreeableness”. This places personality variables in line with other, more biological measures like systolic blood pressure, which also shows greater sex differences in more developed countries (Hottenga et al., 2005).

These findings add to the general pattern known as the “Gender Equality Paradox”: despite societal and political interventions to increase gender equality in many western countries, it is exactly these countries that show the strongest underrepresentation of girls and women in scientific, technological, engineering and mathematics (STEM) professions. Recently, it was shown that more women are represented in STEM fields in countries with less gender egalitarian policies (Stoet & Geary, 2018).

While the robustness of sex differences in personality was demonstrated clearly, studies attempting to explain the correlation of the magnitude of differences with gender equality have suffered from various methodological shortcomings. First, they either did not differentiate between measures for gender equality and overall development, or used measures that confound development and gender equality. While indices for both variables correlate to some degree, this relation is by no means perfect. There are highly developed and wealthy countries with low levels of gender equality, like Saudi-Arabia or South Korea. Conversely, countries with relatively low development can surpass Western countries in terms of gender equality, like Uganda.

Second, many studies have used the five factors model of personality ("Big Five", Goldberg, 1993) to study sex differences. Those five broad domains can be divided into facets or aspects that can in turn show divergent sex differences. For example, the personality domain of “Openness” can be divided into six aspects. Women score higher for the aspects “Openness to feelings” and “Aesthetic interests”, while men score higher for “Openness to ideas” (Weisberg, DeYoung, & Hirsh, 2011). If studied on the broad domain level, these differences would balance each other out.

Third, most studies do not adequately address psychometric challenges in the measurement of personality. It has been argued by individual differences researchers (Del Giudice, Booth, & Irwing, 2012; Giudice, 2009) that personality scores should be studied as latent variables that are estimated under the assumption of measurement invariance. This should ensure that the same constructs are measured in the compared groups. Sex differences are best represented by comparison along multiple variables. For example, differences between men and women in facial morphology are represented by many aspects like facial width, eye size or eyebrow thickness. Human personality is a multidimensional construct as well. While sex differences in personality measures tend to be rather small when considering one facet at a time, these differences can add up to a remarkably large overall difference when taking into account differences among many variables.

This study seeks to tackle all these issues. A large, multi-national dataset of 30 personality facet scores based on the Big Five will be used to study sex differences. Latent scores with measurement invariance constraints will be estimated to get a more accurate representation of these differences. A multivariate effect size measure, Mahalanobis' *D*, will be used to estimate overall sex differences in personality for every country. *D* can be interpreted like Cohen's *d*, as both metrics represent standardized differences between two points of central tendency. However, *D* takes into account differences in several variables and their intercorrelations and provides a measure of distance between two centroids in a multi-dimensional space. In an attempt to replicate findings on the correlation between nations' gender equality indices and sex differences in personality, the Global Gender Gap Index will be related to these difference scores.

Methods

Instrument and Dataset

Measures of personality facets.

The IPIP-NEO is a personality inventory based on the five factor model of personality. It was constructed from the International Personality Item Pool (IPIP) to match the scales of the NEO Personality Inventory (Costa & McCrae, 1992). It is available as a 120-item and a 300-item version. Both versions were found that to correlate highly with those of the NEO-PI, while their reliabilities and validities were shown to be superior to the original (John A. Johnson, 2014). The instrument is based on the five factor model of personality (Big Five). Each of the five broad factor scales comprise six “facet” scales. The five factor scales are Neuroticism (facet scales: Anxiety, Anger, Depression, Self-consciousness, Immoderation and Vulnerability), Extraversion (Friendliness, Gregariousness, Assertiveness, Activity level, Excitement seeking, Cheerfulness), Openness (Artistic interests, Emotionality, Adventurousness, Intellect, Liberalism), Agreeableness (Trust, Morality, Altruism, Cooperation, Modesty, Sympathy) and Conscientiousness (Self-efficacy, Orderliness, Dutifulness, Achievement-striving, Self-discipline and Cautiousness).

Personality Dataset. Two large datasets were acquired from a public repository (Johnson, 2015) that were used for the development of the IPIP-120 (Ns = 307,313 and 619,150). Both samples were collected in an online survey (Johnson, 2014). According to Johnson (2014), these datasets were placed “in the public domain for analyses by interested members of the personality research community”. These datasets contain cases of item responses to the 120 and 300 items version of the IPIP-NEO that were collected in the years 2001 to 2011. Both datasets were combined into one set (N = 926,383), retaining only the items of the 120-item version. Additionally, the dataset contains demographic information on age, sex, country of residence and the date of assessment.

Gender Inequality. Since 2006, the World Economic Forum publishes annual reports called *Global Gender Gap Report*. These reports report on 14 key indicators (e.g. salaries, enrollment in higher education, life expectancy, public office positions) in which men and women differ. Data on these key indicators are aggregated to an overall value of gender inequality: the Global Gender Gap Index (GGGI). This index can theoretically reach from 0 to 1, where 0 would indicate total inequality and 1 total parity. GGGI scores for the years 2006 to 2011 were obtained from publicly available reports at weforum.org. To account for possible changes in GGGI during the assessment period, an average GGGI was calculated from individual years' indices. GGGI scores were obtained for 70 countries for which personality data were available. No GGGI data were available for Antarctica, Serbia, Puerto Rico, Andorra, Taiwan, Afghanistan and Hong Kong.

Per Capita Gross Domestic Product. As an indicator for general economic development of a country, the gross domestic product based on purchasing power parity (GDP-PPP) was obtained from the 2014 World Development Indicators database (United Nations, 2018). The GDP-PPP is the sum of gross value of products in an economy, expressed as international Dollars using purchasing power parity rates and divided by the total population.

Data Analysis

Sample Size Calculation and Resampling. A simulation was conducted to determine the minimum sample size per Country to calculate Mahalanobis' *D*. Assuming a "true" effect of D = 2.71 based on the findings by Del Giudice et al. (2012), A multivariate data set with 40,000 subjects, 30 variables and two groups was simulated. Zero to small mean group differences in variables were simulated to match previous findings on small univariate sex differences in personality facets. Variable intercorrelations were set according to intercorrelations of personality facets in the data set. In 10,000 trials, random samples ranging from 250 to 20000 subjects were drawn and Mahalanobis' *D* was calculated. In less than 95% of trials with a sample size of N > 250, *D* was overestimated by more than .1. Hence, a minimum sample size of 250 subjects was chosen to be the requirement for countries to be included in the analysis. If a country's sample size surpassed 10000 participants, a random sample of 5000 men and 5000 women was drawn from all the country's available subjects. This was done for two reasons. First, the bootstrapping procedure for *D* scores would have been too computationally intensive. Second, as weighted correlation coefficients were to be calculated, those heavily overrepresented countries could have distorted these correlations. Resampling was done for the USA, Canada, UK and Australia.

Latent Personality Variable Estimation. Following the guidelines for studying sex differences in psychological constructs proposed by Del Giudice et al. (2012), latent mean scores were used instead of observed scale scores. These latent scores were estimated using a procedure called multi-group covariance and mean structure analysis (MG-CMSA). In this procedure, several confirmatory factor models with a series of invariance constraints are estimated for scales. The invariance constraints include an equal pattern of factor loadings (configural invariance), the degree of factor loadings (metric invariance) and the intercepts of indicators (scalar invariance). When comparing sex differences, invariance between male and female participants is tested. If the model fit does not decrease significantly while placing additional constraints on the model, measurement invariance is assumed. Cut-off values for assessing model fit were selected according to the guidelines suggested by Hu and Bentler (1999). Models with >.05 for SRMR, >.06 for RMSEA, and ≥.95 for the NNFI and CFI were considered "good fit". When testing measurement invariance, cut-offs were chosen as proposed by Booth and Irwing (2011), who tested invariance for the 16-PF personality inventory. Changes of less than 0.01 for CFI, 0.013 for RMSEA and -.008 for NNFI were considered indicators for invariance.

A random sample (N = 20000) was drawn from the dataset. This sample was used to estimate all factor models. Weighted least squares with robust standard errors ad mean- and variance adjusted test statistics (WLSMV) estimators were used due to the ordinal and skewed nature of the data. Scores for latent constructs were then estimated for the whole sample from the model with scalar invariance.

Univariate mean differences were calculated by computing standardized mean difference scores (Cohen's *d*). The focus of this study, however, was an accurate assessment of overall personality differences along the big five facets. For reasons discussed before, the multivariate generalization of Cohen's *d*, Mahalanobis' *D*, was calculated for mean differences along all 30 facets measured by the IPIP-120. Countries' scores were bootstrapped with 1000 runs per country to obtain 95% confidence intervals (CIs) following the bias-corrected and accelerated method proposed by Kelley (2005).

In addition, two heterogeneity coefficients for *D* scores were calculated: the heterogeneity coefficient H and the "Equal Proportion of Variables" (EPV) value (Del Giudice, 2017). These coefficients indicate whether only a small set of variables contributes to the overall *D* score, or if contributions are evenly distributed among all variables. For the EPV, a cutoff of less than .20 has been proposed to indicate high heterogeneity.

To relate Mahalanobis' *D* scores to the GGGI index, Pearson correlation coefficients were calculated. Unweighted, weighted and GDP-corrected coefficients are reported. Total numbers of participants for each country were used as weights. GDP-corrected coefficients are partial weighted correlations between D and GGGI, controlling for the per capita GDP. Univariate effect sizes for all 30 facets were correlated with the GGGI. This way, the correlation between D and GGGI becomes easier to interpret, as the individual facets' contributions to this overall relationship can be seen.

All of these steps, including simulation, MG-CMSA, further steps of analysis and their results can be reproduced using the R code and data sets provided in the online supplement (Kaiser, 2018).

Results

Fit statistics for all estimated CFA models are summarized in Table 1. Model fit was unsatisfactory for the original model, as both CFI and NNFI did not suggest an acceptable model. Modification indices were calculated and inspected to determine possible changes to model structure. Following the suggestions, cross-loading constraints were partially lifted for a total of 23 items. Mainly, cross-loadings were directed at scales within the same construct. Correlated errors were allowed for two items with similar wording. Details on model structure are given in the supplementary material. The modified model was then tested for measurement invariance. Scalar measurement invariance was achieved with only minimal loss of model fit.

Table 1. Model fit statistics for measurement invariance analysis.

Model	x^2	**df**	**RMSEA**	**SRMR**	**CFI**	**NNFI**
M1: original	332681.96	6585	0.05	0.05	0.91	0.90
M1b: Configural	175100.33	13080	0.04	0.04	0.95	0.95
M2: Metric	179692.83	13212	0.03	0.04	0.95	0.95
Delta M1b-M2	*4592.5*	*132*	*-0.01*	*-0.001*	*-0.001*	*-0.001*
M3: Scalar	182573.83	13302	0.04	0.04	0.95	0.95
Delta M2-M3	*2881*	*90*	*0*	*0*	*0.001*	*-0.001*

Note. M1: original model without cross-loading constraints. M1b: original model with removed constraints on cross loadings and invariance constraints on factor structure. M2: invariance constraints on factor loadings. M3: invariance constraints on intercepts.

Descriptive Results

Participants were 60.02% female and 39.98% male. The mean age of the sample was 25.19 years (SD = 10.15, Median = 21 years). The mean age of female participants was 25.1 years (SD = 10.2, Median = 21 years). For male participants, the mean age was 25.3 years (SD = 10.1, Median = 22 years).

Countries' overall sex difference scores, represented as Mahalanobis' *D*, are summarized in figure 1. *D* scores ranged from 1.48 (Taiwan) to 2.28 (Iceland). The global, sample-size weighted average *D* was 1.84. This corresponds to a 22.03% overlap in distributions between male and female personality scores. Heterogeneity was moderately for most countries: the weighted average H coefficient was 0.756 and the EPV was 0.269. This warranted an investigation of differences in univariate effect sizes for better interpretation of *D* scores, as high heterogeneity indicates that only a part of variables contribute to overall sex differences in personality. Univariate sex differences in personality facets were medium to small, with an averaged absolute Cohen's d of 0.233. To test which facets deviate from zero and contribute most to overall sex differences, CIs were calculated for all 30 individual facets. 95% CIs were Bonferroni-corrected for 30 multiple comparisons, resulting in 99.83% CIs. All results of this analysis are reported in table 2.

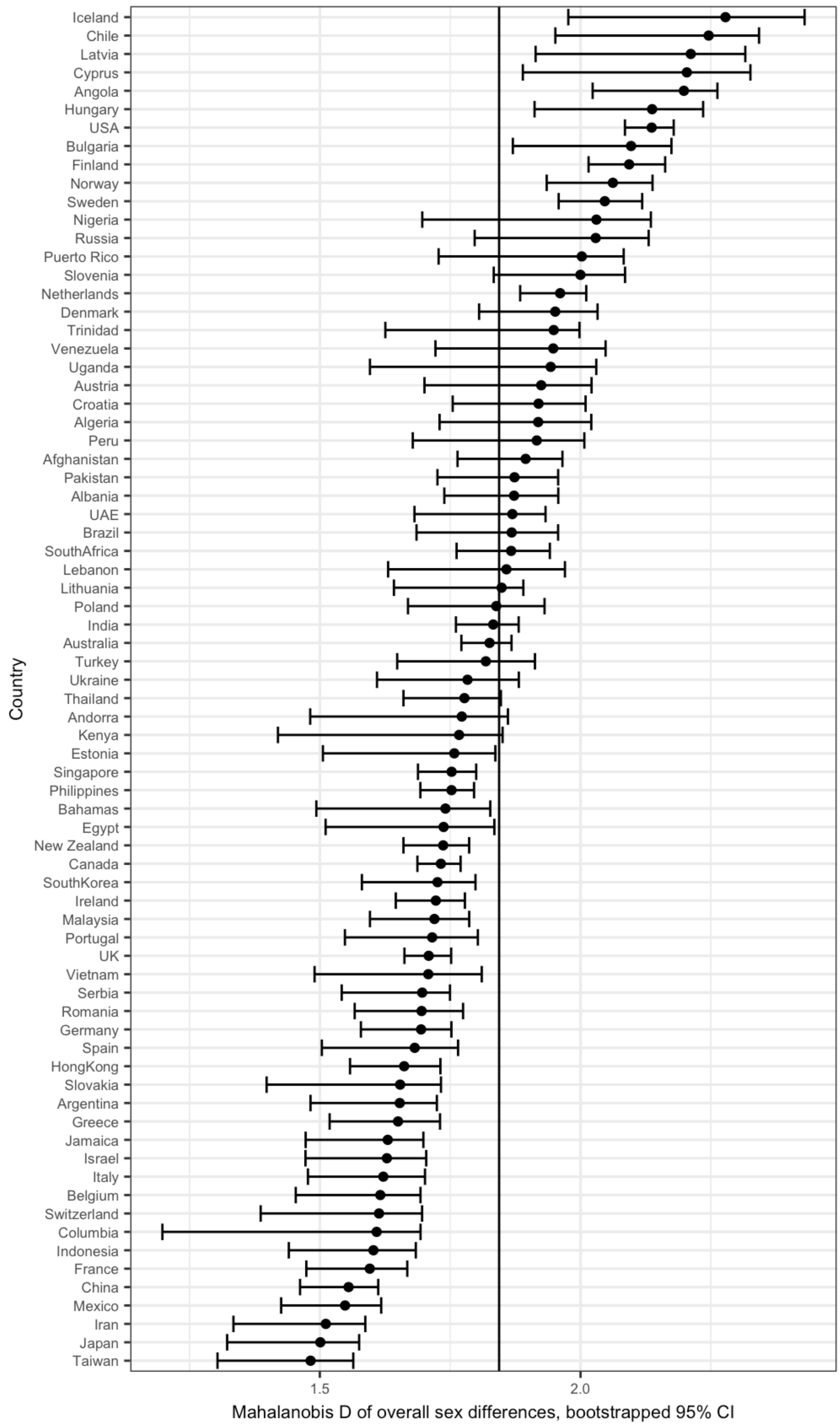
Iceland
Chile
Latvia
Cyprus
Angola
Hungary
USA
Bulgaria
Finland
Norway
Sweden
Nigeria
Russia
Puerto Rico
Slovenia
Netherlands
Denmark
Trinidad
Venezuela
Uganda
Austria
Croatia
Algeria
Peru
Afghanistan
Pakistan
Albania
UAE
Brazil
SouthAfrica
Lebanon
Lithuania
Poland
India
Australia
Turkey
Ukraine
Thailand
Andorra
Kenya
Estonia
Singapore
Philippines
Bahamas
Egypt
New Zealand
Canada
SouthKorea
Ireland
Malaysia
Portugal
UK
Vietnam
Serbia
Romania
Germany
Spain
HongKong
Slovakia
Argentina
Greece
Jamaica
Israel
Italy
Belgium
Switzerland
Columbia
Indonesia
France
China
Mexico
Iran
Japan
Taiwan
Country
1.5
2.0
Mahalanobis D of overall sex differences, bootstrapped 95% CI

Fig. 1. Multivariate sex differences in personality by country and bootstrapped 95% confidence intervals. The vertical line indicates the sample-size weighted global average *D*.

Table 2. Univariate effect sizes for 30 IPIP-120 personality facets, CIs and test statistics of t-tests against a value of 0.

Domain	Facet	*d*	Lower 99.83% CI	Upper 99.83% CI
Agreeableness				
	Trust	-0.01	-0.05	0.04
	Morality	**-0.34**	-0.40	-0.29
	Altruism	**-0.41**	-0.49	-0.33
	Cooperation	**-0.18**	-0.22	-0.13
	Modesty	-0.04	-0.10	0.02
	Sympathy	**-0.54**	-0.60	-0.48
Conscientiousness				
	Self-efficacy	0.04	-0.01	0.09
	Orderliness	-0.02	-0.07	0.03
	Dutifulness	**-0.14**	-0.19	-0.1
	Achievement-striving	**-0.19**	-0.25	-0.14
	Self-discipline	-0.03	-0.08	0.02
	Cautiousness	**0.17**	0.12	0.22
Extraversion				
	Friendliness	**-0.11**	-0.16	-0.07
	Gregariousness	**-0.15**	-0.20	-0.10
	Assertiveness	**0.06**	0.02	0.11
	Activity level	**-0.19**	-0.24	-0.15
	Excitement seeking	0.00	-0.06	0.06
	Cheerfulness	**-0.19**	-0.24	-0.14
Neuroticism				
	Anxiety	**-0.49**	-0.53	-0.44
	Anger	**-0.30**	-0.35	-0.25
	Depression	**-0.14**	-0.18	-0.10
	Self-consciousness	-0.02	-0.07	0.02
	Immoderation	**-0.26**	-0.31	-0.21
	Vulnerability	**-0.47**	-0.52	-0.43
Openness				
	Imagination	**-0.07**	-0.12	-0.01

Artistic interests	**-0.39**	-0.45	-0.34
Emotionality	**-0.69**	-0.76	-0.63
Adventurousness	**0.05**	0.01	0.10
Intellect	**0.23**	0.19	0.28
Liberalism	**-0.13**	-0.17	-0.08

Note. Negative d scores indicate that women score higher on this trait, while positive values indicate higher scores for men. Cohen's d values printed in bold are significantly different from zero based on Bonferroni-corrected CIs.

Relationship of Sex Differences and Gender Equality

GGGI scores of all years were positively correlated with Mahalanobis' *D* values. Statistical significance was achieved for all weighted correlations. GDP-controlled, weighted and unweighted correlation coefficients are reported in table 3. The relation of mean GGGI scores and *D* scores is depicted in figure 2.

Table 3. GDP-controlled, weighted, unweighted and correlations between GGGI scores and Mahalanobis' *D* scores for up to 70 countries, including p values.

	GGGI 2006	**GGGI 2007**	**GGGI 2008**	**GGGI 2009**	**GGGI 2010**	**GGGI 2011**	**GGGI Mean**
Controlled r	.318 (.002)	.381 (.002)	.406 (<.001)	.451 (<.001)	.450 (<.001)	.457 (<.001)	.429 (<.001)
Weighted r	.337 (.006)	.343 (.005)	.374 (.002)	.416 (.001)	.414 (<.001)	.424 (<.001)	.394 (.001)
Unweighted r	.168 (.182)	.169 (.176)	.212 (.088)	.251 (.004)	.230 (.059)	.248 (.041)	.214 (.080)
N	65	67	69	69	70	70	70

Note. Values in round brackets are p values.

To further test the robustness of the main finding, the overall weighted correlation of average gender equality with D was bootstrapped with 10000 runs and bias-corrected and accelerated (BCa) 95% confidence interval was calculated. This resulted in a 95% CI of .054 and .761.

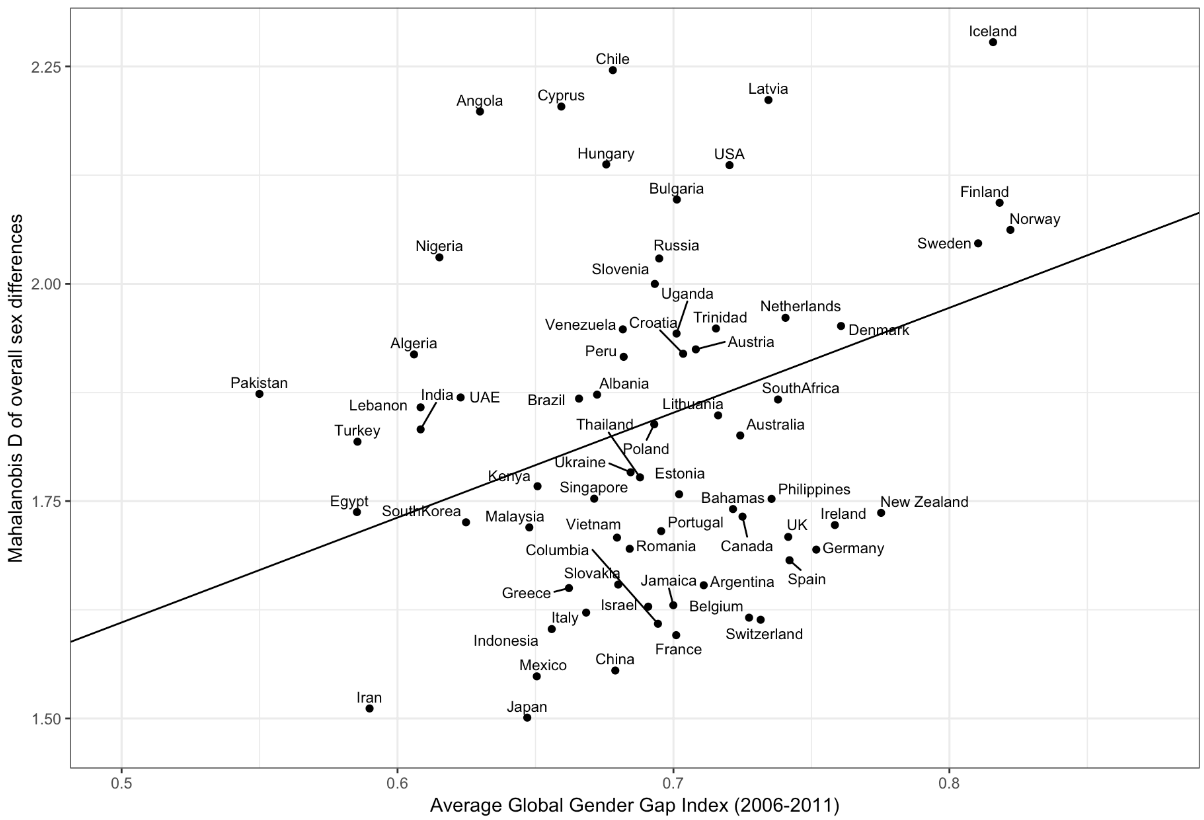

Fig. 2. Scatter plots with best-fitting weighted regression line showing the relation of overall sex difference scores and mean GGGI scores of all countries.

All 30 facets were calculated with the average GGI using Pearson correlations. P values were corrected for multiple comparisons using false discovery rate (FDR) correction. The results are summarized in table 4.

Table 4. Pearson correlations of individual facet differences with the average GGGI and p values.

Domain	Facet	r	p
Agreeableness	Trust	-.24	.05
	Morality	-.16	.18
	Altruism	**-.43**	<.001
	Cooperation	-.04	.74
	Modesty	-.26	.03
	Sympathy	**-0.33**	.01
Conscientiousness	Self-efficacy	-.27	.02
	Orderliness	-.28	.02
	Dutifulness	-.26	.03
	Achievement- striving	**-.45**	<.001
	Self-discipline	-.29	.02
	Cautiousness	.00	.99
Extraversion	Friendliness	**-.32**	.01
	Gregariousness	-.26	.04
	Assertiveness	**-.39**	<.001
	Activity level	**-.29**	.01
	Excitement-seeking	-.21	.09
	Cheerfulness	**-.34**	.01
Neuroticism	Anxiety	.07	.60
	Anger	-.07	.55
	Depression	.06	.65
	Self-consciousness	.27	.03
	Immoderation	**-.35**	<.001
	Vulnerability	.10	.40
Openness	Imagination	.04	.73
	Artistic	-.16	.18
	Emotionality	**-.47**	<.001
	Adventurousness	-.20	.10
	Intellect	.00	1
	Liberalism	-.07	.57

Note. Negative correlations indicate that higher gender equality is associated with women scoring higher compared to men. Correlation coefficients printed in bold are significant after FDR correction.

Discussion

In a large, multinational sample of personality data, the universal nature of gender differences in personality was demonstrated again. Effect sizes of differences were large and comparable to previous studies (Booth & Irwing, 2011; Del Giudice et al., 2012; Giudice, 2009). The relatively high heterogeneity coefficients can be explained by the fact that not all personality facets contribute to the global differences equally. The main contributors to sex differences were mainly related to the domains of agreeableness, negative affect and emotionality. Main contributors to the correlation between sex differences and gender equality were found on all domains.

The overall weighted correlation of gender equality with differences in personality was $r = .394$ ($p < .001$). Following the meta-analytically derived effect size guidelines for studies in differential psychology proposed by Gignac and Szodorai (2016), this can be considered a fairly large effect, as it surpasses 85% of previously found correlations in individual difference research. This finding supports previous work that indicates that sex differences increase with increased gender equality in society. The fact that controlling for GDP did not influence this correlation and sometimes even increased it substantially indicates that this effect can be attributed more directly to gender equality, instead of the state of development of a country.

There are, roughly, two major ways of explaining sex differences in personality. The Social Role Theory of sex differences states that differences in personality variables arise primarily through socialization and the subsequent formation of gender roles (Eagly & Wood, 2016). However, this theory would have predicted smaller sex differences for more gender-egalitarian countries. Social Role Theory thus cannot fully explain cross-cultural variations in sex differences, even though some proportion of differences could still be due to unequal treatment based on sex.

The second explanation, mainly favored by evolutionary psychologists, states that sexual dimorphisms have arisen through differential evolutionary demands on the different sexes. However, evolutionary adaptation processes are expected to take longer to unfold their effects than for gender equality. These two positions can be combined. It can be assumed that differences between the sexes are caused by biological factors as well as environmental influences. The degree to which a society allows individuals to express biological gender differences can vary. If a society ensures that men and women have exactly the same access to all resources that this society has to offer, the biological factors could be expressed more strongly than in more repressive societies. A stronger sexual dimorphism should therefore be seen more as an expression of a successful gender policy. This approach might be appealing, but it does not explain the relatively large contribution of less gender-stereotypical facets to the overall correlation. For example, countries in which women scored higher than men on achievement-striving and assertiveness, gender equality was also higher. Conversely, countries in which men scored higher on self-consciousness than women showed higher equality as well. These findings speak against a purely evolutionary interpretation and could be interpreted as the influence of societal change on human personality.

Limitations

The results presented are limited by the fact that this is a convenience sample of online questionnaires. This distorts the data as the assessment procedures have selected for English language subjects with Internet access. This leads to an over-representation of subjects from English-speaking and more developed countries. In less developed countries, subjects with a higher level of education are likely to be over-represented. Countries with a low level of development are not represented at all because there was not data available or the sample sizes were too small. The risk of calculating distorted difference measurements was too high for them to be included in the analysis.

Since the GGGI was only available for the years 2006 to 2011, only these values could be used for correlative analyses. Nevertheless, some personality data from previous years were included in the analysis. It is quite possible that this has led to an underestimation of the correlation between GGGI and D-scores. Future studies should fill this gap by carrying out larger personality data surveys in less developed countries. It is also important to note that these data cannot be used to draw causal inferences about the relationship between gender equality and personality differences. One way to enable causal inference are large studies on personality carried out in regular intervals so that sufficient data is generated to estimate time-lagged models. This way, future studies could investigate whether changes in personality predict changes in gender equality or vice versa. It could also be that different aspects of personality are subject to different amounts of biological and environmental influences.

Conclusion

Global gender differences in personality are large and can be observed in many different cultures. If anything, policies of gender equality are associated with a greater sexual dimorphism in personality. The results presented suggest that greater sexual dimorphism should not be interpreted as an indicator of a society that discriminates against a particular sex, but rather as an indicator of a successful gender equality policy. This is further supported by the large contributions of facets that are not regarded as sexually dimorphic, like achievement-striving or assertiveness to the overall correlation between sex differences and gender equality.

Future studies should attempt to replicate these results using different data sets and measures, following the general guidelines of using multivariate effect sizes and, if necessary, using latent variable models to estimate psychological variables.

Conflicts of Interest

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no specific funding for this work.

References

Booth, T., & Irwing, P. (2011). Sex differences in the 16PF5, test of measurement invariance and mean differences in the US standardisation sample. *Personality and Individual Differences*, *50*(5), 553–558. https://doi.org/10.1016/j.paid.2010.11.026

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, *4*(1), 5–13. https://doi.org/10.1037/1040-3590.4.1.5

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322–331. https://doi.org/10.1037//0022-3514.81.2.322

Del Giudice, M. (2009). On the Real Magnitude of Psychological Sex Differences. *Evolutionary Psychology*, *7*(2), 147470490900700220. https://doi.org/10.1177/147470490900700209

Del Giudice, M. (2017). Heterogeneity Coefficients for Mahalanobis' D as a Multivariate Effect Size. *Multivariate Behavioral Research*, *52*(2), 216–221.

https://doi.org/10.1080/00273171.2016.1262237

Del Giudice, M., Booth, T., & Irwing, P. (2012). The Distance Between Mars and Venus: Measuring Global Sex Differences in Personality. *PLOS ONE*, *7*(1), e29265. https://doi.org/10.1371/journal.pone.0029265

Eagly, A. H., & Wood, W. (2016). Social Role Theory of Sex Differences. In *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies* (pp. 1–3). American Cancer Society. https://doi.org/10.1002/9781118663219.wbegss183

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, *116*(3), 429.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Goldberg, L. R. (1993). The Structure of Phenotypic Personality Traits. *American Psychologist*, 9.

Hottenga, J.-J., Boomsma, D. I., Kupper, N., Posthuma, D., Snieder, H., & Willemsen, G. (2005). Heritability and Stability of Resting Blood Pressure, 10.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, *51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Johnson, J. A. (2015, September 22). Johnson's IPIP-NEO data repository. Retrieved

February 26, 2018, from https://osf.io/tbmh5/

Kaiser, T. (2018, February 24). Sex Differences in Big Five Personality Facets. Retrieved July 25, 2018, from https://osf.io/6nuhx/

Kelley, K. (2005). The Effects of Nonnormal Distributions on Confidence Intervals Around the Standardized Mean Difference: Bootstrap and Parametric Confidence Intervals. *Educational and Psychological Measurement*, *65*(1), 51–69. https://doi.org/10.1177/0013164404264850

Lippa, R. A. (2010). Sex Differences in Personality Traits and Gender-Related Occupational Preferences across 53 Nations: Testing Evolutionary and Social-Environmental Theories. *Archives of Sexual Behavior*, *39*(3), 619–636. https://doi.org/10.1007/s10508-008-9380-7

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2009). "Why can't a man be more like a woman? Sex differences in big five personality traits across 55 cultures": Correction to Schmitt et al. (2008). *Journal of Personality and Social Psychology*, *96*(1), 118–118. https://doi.org/10.1037/a0014651

Stoet, G. (2018). The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education. *Psychological Science*, 0956797617741719. https://doi.org/10.1177/0956797617741719

United Nations. (2018). GDP per capita (2011 PPP $) | Human Development Reports. Retrieved June 30, 2018, from http://hdr.undp.org/en/content/gdp-per-capita-2011-ppp

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender Differences in Personality across the Ten Aspects of the Big Five. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00178